Noname manuscript No.
(will be inserted by the editor)

Weighted likelihood latent class linear regression

Received: date / Accepted: date

Abstract A weighted likelihood approach for robust fitting a finite mixture of linear regression models is proposed. An EM type algorithm and its variant based on the classification likelihood have been developed. The proposed algorithm is characterized by an M-step that is enhanced by the computation of weights aimed at downweighting outliers. The weights are based on the Pearson residuals stemming from the assumption of normality for the error distribution. Formal rules for robust clustering and outlier detection are also defined based on the fitted mixture model. The behavior of the proposed methodologies has been investigated by some numerical studies and real data examples in terms of both fitting and classification accuracy and outlier detection.

Keywords Classification · EM · Mixture · Outliers detection · Pearson residuals · Regression · Robustness · Weighted Likelihood

Mathematics Subject Classification (2000) MSC 62F35 · MSC 62G35 · MSC 62H25 · MSC 62H30

1 Introduction

The problem of clustering around linear structures is particularly appealing and has received growing interest in the literature. Latent class regression has applications in many fields, including engineering, genetics, biology, econometrics, marketing, computer vision, pattern recognition, tomography, fault detection, among others. The reader is pointed to García-Escudero et al (2009) for a large collection of references. This paper is motivated by the fact that noisy data frequently appear in every field of application. When the sample data is contaminated by the occurrence of outliers, it is well known that maximum likelihood estimation (MLE) is likely to lead to unreliable results. In a mixture setting, the bias of at least one of the component parameters estimate can be arbitrarily large and the true underlying clustering structure can be hidden. Therefore, there is the need for a suitably robust procedure providing protection against outliers. The reader

Address(es) of author(s) should be given

is pointed to the book by Farcomeni and Greco (2015a) for a gentle introduction to robustness issues.

The problem of robust fitting of a mixture of linear regressions has been already tackled in the literature. In general, the robust solutions are driven by a suitable modification of the EM algorithm for mixtures or the classification EM algorithm (CEM), concerning the M step, which is enhanced by some robust estimation approach in place of maximum likelihood. Some existing proposals are based on the idea of (hard) trimming: estimation is performed over a subset of the original data obtained after discarding those units with the lowest contributions to the likelihood function. According to such trimming strategies, potential outliers are discarded in the estimation process, that is observations are given crispy weights in $\{0, 1\}$. Neykov et al (2007) introduced a mixture fitting approach based on the trimmed likelihood, García-Escudero et al (2010) extended the TCLUST methodology, developed in García-Escudero et al (2008) for mixtures of multivariate Gaussian distributions, exploiting the idea of impartial trimming in TCLUST-REG, a related proposal has been presented in García-Escudero et al (2009) and an adaptive hard trimming procedure has been described in Riani et al (2008) based on the Forward Search methodology. In particular, TCLUST-REG is characterized by group scatter constraints aimed at making the mixture fitting a well-posed problem and the addition of a second trimming step to mitigate the effect of outliers in the space of explanatory variables acting as leverage points. A very recent adaptive version of TCLUST-REG has been discussed in Torti et al (2019). An alternative approach meant to automatically take into account leverage points has been considered by García-Escudero et al (2017) where trimming and restrictions have been introduced to get a robust version of the cluster weighted model, named Trimmed Clustered Weighted Restricted Model (TCWRM). In this approach restrictions concern both the set of eigenvalues of the covarance matrix evaluated on the $X-$space and the variances of the regression error term. The reader is pointed to Torti et al (2019) for a comparative analysis of TCLUST-REG and TCWRM under general settings. The benefits of trimming for robust regression clustering have been also investigated in Dotto et al (2017) where a fuzzy approach has been developed.

In a different but complementary fashion, Bashir and Carter (2012) and Bai et al (2012) modified the M step by resorting to soft rather than hard trimming procedures. Actually, they replaced the single component MLE problems by M- (and S-) estimation problems for linear regression (see also Campbell (1984) and Maronna et al (2018)). In particular, in both papers the authors developed an EM-type algorithm featured by componentwise weights but this approach can be extend to obtain robust versions of the CEM algorithm based on M- and S-estimation, as well. According to a soft trimming strategy, observations are attached a weight lying in $[0, 1]$ according to some measure of outlyingness. Potential outliers are expected to be heavily downweighted, whereas genuine observations receive a weight close to one.

It is worth to mention that there are different proposals aimed at robust latent class linear regression estimation that are not based on soft or hard trimming procedures in which the assumed model is embedded in a larger one to account for outliers. Yao et al (2014) considered a mixtures of linear regression models with Student t error distributions; Punzo and McNicholas (2017) developed an approach based on the Contaminated Gaussian Cluster Weighted Model in which

each mixture component has some parameters controlling the proportion of (different type of) outliers; Yu et al (2017) proposed a case-specific and scale-dependent mean-shift mixture model and a penalized likelihood approach to induce sparsity among the mean-shift parameters.

Here, we propose the use of the weighted likelihood methodology (Markatou et al, 1998) as a valid alternative to the existing methods. Weighted likelihood is an appealing robust techniques for estimation and testing (Agostinelli and Markatou, 2001). In particular, reliable statistical tools have been developed for linear regression (Agostinelli and Markatou, 1998; Agostinelli, 2002), generalized linear models (Alqallaf and Agostinelli, 2016) and multivariate analysis (Agostinelli and Greco, 2019). Recently, Greco and Agostinelli (2019) also introduced weighted likelihood estimation of mixtures of multivariate normal distributions. The authors explored the behavior of both EM and CEM type algorithms and found that weighted likelihood gives powerful devices for robust estimation, classification and outliers detection. Then, the same ideas can be extended to the context of mixtures of linear regressions.

Weighted likelihood belongs to the group of soft trimming techniques and the weighted likelihood estimator (WLE) can be thought as an M-estimator. The main differences are in the genesis of the weights and in their asymptotic behavior at the assumed model. Actually, weighted likelihood estimation can correspond to a minimum disparity estimation problem (Basu and Lindsay, 1994). Then, conversely to M-estimators, the WLE is asymptotically efficient at the model and is expected to be highly robust under contamination. Some necessary preliminaries on weighted likelihood estimation are given in Section 2. The weighted EM and penalized CEM algorithms for robust fitting of mixtures of regressions are introduced in Section 3, while outlier detection rules are outlined in Section 4. Some illustrative examples based on simulated data are presented in Section 5 and Section 6 gives some numerical studies. A real data example is discussed in Section 7.

2 Background

Let $y = (y_1, \cdots, y_n)^{\mathrm{T}}$ be a random sample of size n drawn from a r.v. Y with distribution function $M(y;\theta)$ and probability (density) function $m(y;\theta)$, which is an element of the parametric family of distributions $\mathcal{M} = \{M(y;\theta), \theta \in \Theta \subseteq \mathbb{R}^d, d \geq 1, y \in \mathcal{Y}\}$. Let $\hat{F}$ be the empirical distribution function. The WLE $\hat{\theta}^w$ is defined as the root of the Weighted Likelihood Estimating Equations (WLEE)

$$\sum_{i=1}^{n} w(y_i;\theta,\hat{F}) s(y_i;\theta) = 0 \ , \tag{1}$$

where $s(y;\theta) = \sum_{i=1}^{n} s(y_i;\theta)$ is the score function. The WLEE in (1) is a modified version of the (system of) likelihood equations, since a data dependent weight, $w_i = w(y_i;\theta,\hat{F}) \in [0,1]$, is attached to each individual score component. The weights are meant to be small for those data points that are in disagreement with the assumed sampling model. The degree of agreement between the data and the assumed model is measured by the Pearson residual function. Let

$$f^*(y) = \int_{\mathcal{Y}} k(y;t,h) d\hat{F}(t)$$

be a non parametric kernel density estimate and

$$m^*(y;\theta) = \int_{\mathcal{Y}} k(y;t,h)m(t;\theta)\ dt$$

a smoothed version of the model density obtained by using the same kernel function. Then, the Pearson residual is

$$\delta(y) = \frac{f^*(y) - m^*(y;\theta)}{m^*(y;\theta)}\ ,$$

with $\delta(y) \in [-1, +\infty]$. By smoothing the model, the Pearson residuals converge to zero with probability one for every y under the assumed model; the reader is pointed to Basu and Lindsay (1994), Markatou et al (1998) and references therein. When the model is discrete, $f^*(y)$ is the empirical probability function and $m^*(y;\theta)$ simply reduces to $m(y;\theta)$. In this paper, we will make use of the Pearson residuals established in Agostinelli and Greco (2019). Actually, a valid WLEE can be also obtained by using Pearson residuals that are defined as

$$\delta(y) = \frac{f^*(\tilde{y}) - m^*(\tilde{y})}{m^*(\tilde{y})}\ ,$$

where $\tilde{y} = g(y;\theta)$ is a pivot at the assumed model whose (smoothed) distribution does not depend on the parameter value.

Large values of the Pearson residual function correspond to regions of the support of Y where the model fits the data poorly. According to this approach, outliers can be defined as *observations that are highly unlikely to occur under the assumed model*, rather than from a geometric point of view as observation that are far from the model fitted to the bulk of the data, as in the classical theory of M-estimators.

The weight function is defined as

$$w(\delta(y)) = \frac{[A(\delta(y)) + 1]^+}{\delta(y) + 1}\ , \tag{2}$$

where $[\cdot]^+$ denotes the positive part and $A(\delta)$ is the Residual Adjustment Function (RAF, Basu and Lindsay (1994)). The RAF plays the role to bound the effect of large Pearson residuals on the fitting procedure. By using a RAF such that $|A(\delta)| \leq |\delta|$ both outliers and inliers (whose nature will be described in the following) will be downweighted. The RAF function is connected to minimum disparity estimation problems. Actually, it is defined as $A(\delta) = (\delta+1)G'(\delta) - G(0)$, with prime denoting differentiation, where $G(\cdot)$ is a strictly convex function over $[-1, \infty]$ and thrice differentiable, which determines a disparity measure, that, in the continuous case, is defined as

$$\rho(f^*(y), m^*(y;\theta) = \int_{\mathcal{Y}} G(y)m^*(y;\theta)\ dy\ .$$

In principle, by following the approach developed in Markatou et al (1998), it is possible to build a WLEE matching a minimum disparity objective function. One can consider the families of RAF stemming from the Symmetric Chi-Squared divergence, the family of Power divergence or Generalized Kullback-Leibler divergence measures. The resulting weight function is unimodal and decline smoothly

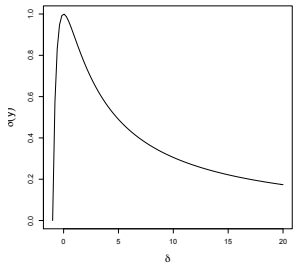

Fig. 1 Weighting function corresponding to a Symmetric Chi-Squared divergence

to zero as $\delta(y) \to -1$ or $\delta(y) \to \infty$. The weighting function corresponding to a Symmetric Chi-Squared divergence, which is driven by $G(\delta) = \frac{2\delta^2}{\delta+2}$, is given in Figure 1.

Under the assumptions given in Markatou et al (1998) and Agostinelli and Markatou (2001), that establish some regularity conditions on the model, the kernel and the weight function, at the assumed model, we have that:

1. $\hat{\theta}^w$ is a consistent and first order efficient estimator of θ, that is

$$\sqrt{n}(\hat{\theta}^w - \theta) \xrightarrow{d} N(0, I_1^{-1}(\theta))$$

   where $I_1(\theta) = E[u(Y;\theta)^2]$ is the expected Fisher information;
2. $\sup|w(y, \hat{\theta}^w, \hat{F}) - 1| \xrightarrow{a.s.} 0$ (Agostinelli and Greco, 2013);
3. the weighted versions of the likelihood ratio, Wald and score test all share the usual asymptotic behavior (Agostinelli and Markatou, 2001).

It is worth to claim that the shape of the kernel function has a very limited effect on weighted likelihood estimation. On the contrary, the smoothing parameter h allows to control the robustness/efficiency trade-off of the methodology in finite samples. Actually, large values of h lead to Pearson residuals all close to zero and weights all close to one and, hence, large efficiency, since the kernel density estimate is stochastically close to the postulated model. On the other hand, small values of h make the kernel density estimate more sensitive to the occurrence of outliers and the Pearson residuals become large for those data points that are in disagreement with the model.

2.1 Weighted likelihood for linear regression

Let us consider a linear regression model with normally distributed errors, i.e. $y = X\beta + \sigma\epsilon$, where y is a response variable, $X = [x_1, \ldots, x_p]$ is the $n \times p$ design matrix, $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the vector of regression coefficients, σ is a scale parameter and $\epsilon \sim N(0,1)$. In this setting, Pearson residuals and the weights can be evaluated over the scaled residuals $g(y;\beta,\sigma) = e = (y - X\beta)/\sigma$. An appealing strategy to compute Pearson residuals consists in using a normal kernel with bandwidth

equal to h. In such a way, the smoothed model density is still normal with variance $(1+h^2)$, that is

$$\delta(y) = \frac{f^*(e)}{\frac{1}{\sqrt{1+h^2}}\phi\left(\frac{e}{\sqrt{1+h^2}}\right)} - 1 \ , \tag{3}$$

where $\phi(\cdot)$ denotes the standard normal density function. Then, the WLE of (β, σ) is obtained as the result of a weighted least squares. Clearly, the computation of the WLE of (β, σ) yields an iterative procedure. At each iteration, based on the current parameter estimates, scaled residuals are obtained. Then, their non parametric density estimate is fitted based on the chosen kernel and Pearson residuals and weights are updated according to (3) and (2).

3 Robust fitting of a latent class linear regression model

Let us assume a latent class regression model featured by K components, where K is fixed in advance, with density function denoted by

$$m(y; x, \tau) = \sum_{k=1}^{K} \pi_k \phi(y; \mu_k, \sigma_k), \tag{4}$$

where $\mu_k = X\beta_k$, π_k is the prior probability of component k, (β_k, σ_k) are the component specific parameters and $\tau = (\pi_1, \ldots, \pi_K, \beta_1, \ldots, \beta_K, \sigma_1, \ldots, \sigma_K)^{\mathrm{T}}$ is the vector of all parameters.

The mixture loglikelihood function based on a sample of size n is

$$\ell(\tau) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \pi_k \phi(y_i; \mu_k, \sigma_k) \ .$$

Maximum likelihood estimation is commonly performed by the EM algorithm, that works with the classification loglikelihood

$$\ell_c(\tau) = \sum_{i=1}^{n} \sum_{k=1}^{K} \log(\pi_k \phi(y_i; \mu_k, \sigma_k)) u_{ik} \ ,$$

where u_{ij} is an indicator of the ith unit belonging to the jth cluster. The EM algorithm iterates, over the index s, between the E step, in which posterior membership probabilities are evaluated as

$$u_{ik}^{(s)} = \frac{\pi_k^{(s-1)} \phi\left(y_i; \mu_k^{(s-1)}, \sigma_k^{(s-1)}\right)}{\sum_{k=1}^{K} \pi_k^{(s-1)} \phi\left(y_i; \mu_k^{(s-1)}, \sigma_k^{(s-1)}\right)}$$

and the M step, where parameters' estimates are updated as

$$\begin{aligned}
\pi_k^{(s)} &= \frac{\sum_{i=1}^{n} u_{ik}^{(s)}}{n} \\
\beta_k^{(s)} &= (X^{\mathrm{T}} U^{(s)} X)^{-1} X^{\mathrm{T}} U^{(s)} y \\
\sigma_k^{2(s)} &= \frac{\left(y - \mu_k^{(s)}\right)^{\mathrm{T}} U^{(s)} \left(y - \mu_k^{(s)}\right)}{n}
\end{aligned}$$

where $U^{(s)}$ is the $n \times k$ matrix whose i-th row is $u_{i.}^{(s)} = \left(u_{i1}^{(s)}, \ldots, u_{iK}^{(s)}\right)$. At convergence, cluster assignments can be pursued according to a Maximum a Posteriori (MAP) rule: units are assigned to the most likely component. In the CEM algorithm, after the E step, a classification step is performed (together they form the CE step). Let $k_i = \text{argmax}_k u_{ik}^{(s)}$, then $u_{ik_i}^{(s)} = 1$ and $u_{ik}^{(s)} = 0$ for $k \neq k_i$ and $U^{(s)}$ becomes a dummy matrix. Conversely to the EM algorithm, the CEM directly provides a classification of the units at convergence. Actually, the classification approach is aimed at maximizing the classification loglikelihood over both the mixture parameters and the individual components' labels.

Weighted versions of the above algorithms can be designed by introducing the computation of the weights defined in (2) before the M step at the current parameter value. In particular, the weighted EM (WEM) will require componentwise sets of weights, wheres in the weighted CEM (WCEM) weights will be computed conditionally on the current cluster assignments driven by the CE step. More in details, the WEM algorithm iterates between the classical E step and an M step in which the single components MLE problems are replaced by one-step WLE problems. The single iteration is summarized in Algorithm 1.On the contrary, the WCEM algorithm iterates between the standard CE step and a one-step weighted likelihood based M-step in which weights are evaluated conditionally to the current cluster assignment, that is $w_{ik} = w_{ik_i}$ and not for each component anymore.

3.1 Computational details

One of the first issues to deal with the estimation of a mixture model by the EM or CEM algorithm and their robust counterparts is the choice of a suitable starting point. A solution is represented by subsampling (Markatou et al, 1998; Neykov and Müller, 2003; Neykov et al, 2007; Torti et al, 2019). A subsample of size n^* is selected randomly from the data sample, then the model is fitted to these n^* observations by the classical EM (or CEM) algorithm to get a trial estimate. In order to avoid the algorithm to be dependent on initial values, a simple and common strategy is to run the algorithm from a number of starting values. This approach shows some limitations since from the one hand n^* should be as small as possible in order to increase the chance of drawing at least one outlier free subsample, but from the other hand a larger trial sample size will avoid the algorithm to fail in finding a solution.

Here, in a different fashion, a deterministic initialization will be considered: first units are assigned to the different components by running TCLUST to the multivariate data (y, X), then cluster specific parameters' estimates are initialized by running a robust regression conditionally on clusters' assignments. In particular, weighted likelihood regression has been used but M-type regression could be used as well. The initial clustering depends on a couple of tuning constants that allow control of TCLUST: the level of trimming α and an eigen-ratio constraint factor c (García-Escudero et al, 2008; Fritz et al, 2013). A general advise is to run the algorithm few times for different values of (α, c). This strategy is well justified since in García-Escudero et al (2010) it is stated that TCLUST could serve as starting point for others approaches.

An alternative deterministic initial solution may be obtained by computing the trimmed likelihood estimator of Neykov et al (2007); other candidate initial solu-

Algorithm 1 Computation of weights and the M step of the WEM algorithm

Weights
for $c = 1, \ldots, k$ **do**

$$e_{ik}^{(s)} = \frac{1}{\hat{\sigma}^{(s)}}(y_i - \hat{\mu}_k^{(s)})$$
$$\delta_{ik}^{(s)} = \frac{f^*\left(e_{ik}^{(s)}\right)}{\phi(e_{ik}^{(s)}; 0, \sqrt{(1+h^2)})} - 1$$

Obtain

$$w_{ik}^{(s)} = \frac{\left[A\left(\delta_{ik}^{(s)}\right) + 1\right]^+}{\delta_{ik}^{(s)} + 1}$$

end for
M-step
for $c = 1, \ldots, k$ **do**

$$\pi_k^{(s+1)} = \frac{\sum_{i=1}^{n} u_{ik}^{(s)} w_{ik}^{(s)}}{\sum_{i=1}^{n}\sum_{k=1}^{K} u_{ik}^{(s)} w_{ik}^{(s)}}$$
$$\beta_k^{(s+1)} = (X^{\mathrm{T}} \tilde{W}^{(s)} X)^{-1} X^{\mathrm{T}} \tilde{W}^{(s)} y$$
$$\sigma_k^{(s+1)} = \frac{\sum_{i=1}^{n}\left(y_i - \mu_k^{(s+1)}\right)^2 w_{ik}^{(s)} u_{ik}^{(s)}}{\sum_{i=1}^{n} w_{ik}^{(s)} u_{ik}^{(s)}},\ \mu_k = X\beta_k$$

with $\tilde{W}^{(s)} = \left[\tilde{w}_{ik}^{(s)}\right]$, with $\tilde{w}_{ik} = u_{ik} w_{ik}$
end for

tions can be evaluated according to the approach discussed in Coretto and Hennig (2017) that is based on a combination of nearest neighbor denoising and agglomerative hierarchical clustering. Further starting points can be obtained by randomly perturbing the deterministic starting solution and/or the final one obtained from it (Farcomeni and Greco, 2015b).

Bad staring point can lead to spurious solutions, characterized by an excess of downweighting, or to even non robust solutions. Some guidance for root selection can be provided by the sum of the weights at convergence. The weights are evaluated conditionally on the final cluster assignments, that is $\hat{w}_i = \hat{w}_{ik_i}$. Actually, if $\sum_{i=1}^{n} \hat{w}_i \approx 1$, the WLE is close to the MLE, whereas if $\sum_{i=1}^{n} \hat{w}_i$ is too small, then the corresponding WLE is a degenerate solution, indicating that it only represents a small subset of the data. Therefore, the monitoring of a summary of the weights at convergence as the initialization varies is a suitable strategy to detect eventual different solutions for a fixed h, that all need to be explored in order to catch the main features of the data. In particular, it is of interest to monitor the empirical downweighting level $(1 - \hat{\bar{\omega}})$, with $\hat{\bar{\omega}} = n^{-1} \sum_{i=1}^{n} \hat{w}_i$, which can be interpreted as a rough approximation of the level of contamination in the data according to the fitted model.

A further adaptive strategy is that of selecting the most frequent root (the modal root in the terminology adopted by Bai et al (2012)) when several roots are found by running the algorithm from several starting points.

Formal solutions to the problem of root selection in weighted likelihood estimation have been provided in Markatou et al (1998); Agostinelli (2006); Agostinelli and Greco (2019). Here, we decided to select the root leading to the minimum fitted approximate disparity as defined in Agostinelli and Greco (2019), that is

$$\tilde{\rho}(f^*, m^*) = \frac{1}{n} \sum_{i=1}^{n} \frac{G(\delta_i) + \delta_i}{\delta_i + 1} \tag{5}$$

where the Pearson residuals δ_i are evaluated conditionally on the final cluster assignments, that is $\delta_i = \delta_{ik_i}$, at convergence.

Another crucial aspect is represented by the selection of the bandwidth parameter h. The tuning of the smoothing parameter h could be based on several quantities of interest stemming from the fitted mixture model: a safe selection can be achieved by monitoring the unit specific weights, residuals or the empirical downweighting level as h varies (Markatou et al, 1998; Greco, 2017; Agostinelli and Greco, 2017). An abrupt change in the monitored empirical downweighting level or in the residuals from the robust fit may indicate the transition from a robust to a non robust fit and aid in the selection of a value of h that gives an appropriate compromise between efficiency and robustness at finite samples. A monitoring approach is commonly applied to select the trimming level in TCLUST, TCLUST-REG and TCWRM, for instance. The performance of monitoring and the strategy based on the criterion (5) will be illustrated in the following sections. The reader is pointed to Cerioli et al (2018) for a recent general account on the benefits and potentials of monitoring.

In addition, it is worth mention that the proposed algorithm can be successfully augmented by introducing scatter similarity restrictions as described by García-Escudero et al (2010). These constraints are posed by fixing a constant c such that

$$\frac{\max \sigma_k}{\min \sigma_k} \leq c, \ k = 1, 2, \ldots, K$$

and are needed to avoid spurious solutions and make the mixture fitting and classification well defined problems (see also Fritz et al (2013); Garcia-Escudero et al (2015); Greco and Agostinelli (2019)).

3.2 Properties

The WEM and WCEM are obtained by replacing maximum likelihood by a different set of estimating equations, characterized by the introduction of weights aimed at bounding the effect of outliers on the fit. In a fashion similar to what stated in Bai et al (2012), the proposed algorithms represent a special case of the algorithm first introduced by Elashoff and Ryan (2004), where an EM algorithm has been established for very general estimating equations. Here, in the M-step, it is suggested to solve a complete data estimating equation of the form

$$\Psi(y; X, \tau) = (\Psi_\pi(y; X, \tau), \Psi_\beta(y; X, \tau), \Psi_\sigma(y; X, \tau))^{\mathrm{T}} = 0 \tag{6}$$

with

$$\Psi_\pi(y;X,\tau) = (\Psi_{\pi_1}(y;X,\tau),\ldots,\Psi_{\pi_K}(y;X,\tau))^{\mathrm{T}},$$
$$\Psi_\beta(y;X,\tau) = (\Psi_{\beta_1}(y;X,\tau),\ldots,\Psi_{\beta_K}(y;X,\tau))^{\mathrm{T}},$$
$$\Psi_\sigma(y;X,\tau) = (\Psi_{\sigma_1}(y;X,\tau),\ldots,\Psi_{\sigma_K}(y;X,\tau))^{\mathrm{T}}$$

and

$$\Psi_\pi(y;X,\tau) = \sum_{i=1}^{n} \psi_{\pi_j}(y_i;\tau)u_{ij} = \sum_{i=1}^{n} w(y_i;\tau,\hat{F})s_{\pi_j}(y_i;\tau)u_{ij},$$
$$\Psi_\beta(y;X,\tau) = \sum_{i=1}^{n} \psi_{\beta_j}(y_i;\tau)u_{ij} = \sum_{i=1}^{n} w(y_i;\tau,\hat{F})s_{\beta_j}(y_i;\tau)u_{ij},$$
$$\Psi_\sigma(y;X,\tau) = \sum_{i=1}^{n} \psi_{\sigma_j}(y_i;\tau)u_{ij} = \sum_{i=1}^{n} w(y_i;\tau,\hat{F})s_{\sigma_j}(y_i;\tau)u_{ij}.$$

Very general conditions for consistency and asymptotic normality of the solution to (6) are given in Elashoff and Ryan (2004), whereas Bai et al (2012) gives conditions in the case of M-estimators. The main requirements are that

1. ψ defines an unbiased estimating function, i.e. $E_\tau[\psi(Y;X,\tau)] = 0$;
2. $E_\tau[\Psi(Y;X,\tau)\Psi(Y;X,\tau)^{\mathrm{T}}]$ exists and is positive definite;
3. $E_\tau[\partial\Psi(Y;X,\tau)/\partial\tau]$ exists and is negative definite, $\forall\tau$.

This conditions are satisfied by the proposed WLEE, that are characterized by weighted score functions as in (6) (see also the Supplementary material in Agostinelli and Greco (2019)). Since the WLEE can be considered as M-type estimating equations and all the above requirements are fulfilled, one can state the following result, along the lines of Bai et al (2012). Under the regularity conditions of Section 2, under the further identifiability conditions of the model (4) given in Hennig (2000), existence, consistency and asymptotic normality of the WLE $\hat{\tau}^w$ implicitly defined by equation (6) hold. In particular, the asymptotic covariance matrix of $\hat{\tau}^w$ can be obtained in the usual sandwich fashion. Consistency is defined conditionally on the true labels and concerns the case in which the WLEE admits a unique solution. Actually, as introduced in the previous subsection, the WLEE may admit multiple roots. The selection of the consistent root can be effectively pursued according to the strategies described in Subsection 3.1.

4 Outlier detection

The WEM and WCEM algorithms lead to classify all the sample units, both genuine and contaminated observations, meaning that also outliers are assigned to a cluster. Actually, we are not interested in classifying outliers and for purely clustering purposes outliers have to be discarded. Outlier detection should be based on the robust fitted model and performed separately by using formal rules. The key ingredients in outlier detection are the (scaled) residuals. For a fixed significance level α, an observation is flagged as an outlier when the corresponding residual in absolute value exceeds a fixed threshold, corresponding to the $(1 -$

$\alpha/2)-$level quantile of the reference standard normal distribution. In the case of finite mixtures, the main idea is that the outlyingness of each data point should be measured conditionally on the final assignment (Greco and Agostinelli, 2019), i.e. an observation is flagged as outlying when

$$\frac{|y_i - X_i\hat{\beta}_{k_i}|}{\hat{\sigma}_{k_i}} > z_{1-\frac{\alpha}{2}} \tag{7}$$

Popular choices are $\alpha = 0.05$ and $\alpha = 0.01$. The process of outlier detection may result in type-I and type-II errors. In the former case, a genuine observation is wrongly flagged as outlier (swamping), in the latter case, a true outlier is not identified (masking). Swamped genuine observations are false positives, whereas masked outliers are false negatives. A measure of the level of the test is provided by the rate of false positives, whereas the power of the testing procedure is given by the rate of true positives. The outliers detection process could also be designed to take into account multiplicity arguments in the simultaneous testing of all the n data points. For instance, one could base the outlier detection rule on the False Discovery Rate (FDR, Cerioli and Farcomeni (2011)).

5 Illustrative examples with synthetic data

The overall behavior of WEM and WCEM is illustrated in the following examples based on simulated data. The proposed methodology has been tested on some data configurations that have been already used in the literature concerning robust fitting of mixtures of regression lines. The interest lies on both fitting and classification accuracy and in the outlier detection testing rule. The WLEE are based on a symmetric Chi-squared RAF. For each example, we display the data with their original clustering and the true regression lines superimposed and, in separated panels, the results stemming from WEM and WCEM. The outlier detection rule relies on the FDR at a 5% level. We use different symbols and colors for the clusters with a black + standing for the detected outliers (and the true outliers in the panel with the true assignments). In every situation the classical EM and CEM algorithms give unreliable results because of contamination in the sample at hand.

Example 1.

Let us consider a mixture of three simple normal linear regressions. The regression lines were generated according to the models

$$\begin{cases} y_1 = 3 + 1.4x + 0.1\epsilon \\ y_2 = 3 - 1.1x + 0.1\epsilon \\ y_3 = 0.2x + 0.1\epsilon \end{cases}$$

with $\epsilon \sim N(0, 1)$ (Neykov et al, 2007). The clusters' sizes are 70, 70, 60, respectively. Then 50 outliers were added that are uniformly distributed in the rectangle that contains the genuine data points. Outliers are such that their distance from the true regression lines, as measured by the scaled residual in absolute value, is above the 0.95-level quantile of the standard normal distribution. The data, the

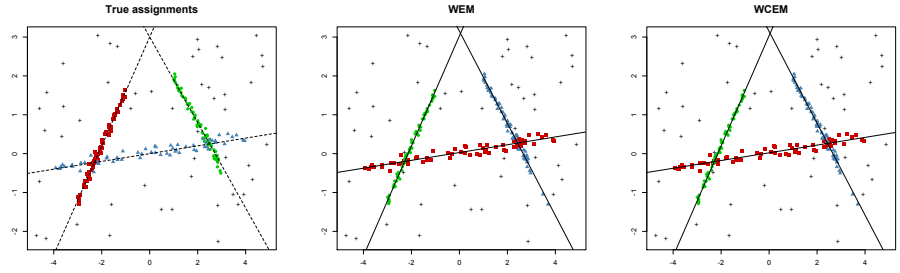

Fig. 2 Example 1. True assignments (left), WEM (middle), WCEM (right).

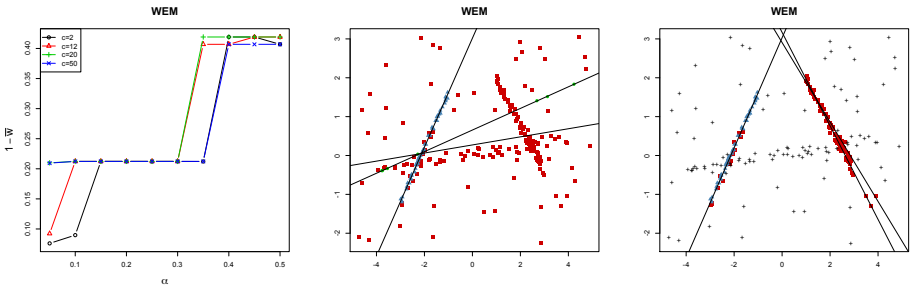

Fig. 3 Example 1. Monitoring of $1 - \bar{w}$ by varying (α, c) of the intial TCLUST (left), WEM root 2 (middle) and WEM root 3 (right).

fitted models and the final classification are displayed in Figure 2: the left panel gives the true assignments and the true lines, the middle panel and the right panel display the results stemming from WEM and WCEM, respectively. The weighted likelihood methodology provides quite satisfactory outcomes both in terms of fitting and classification accuracy.

To illustrate the problem concerning the initialization of the WEM and WCEM algorithms and the selection of the best root, we consider different starting points obtained by varying the tuning parameters of TCLUST (α, c), for a fixed h. The left panel of Figure 3 displays the empirical downweighting level at convergence stemming from WEM. Three different solutions are apparent: in the central part we find the majority of solutions leading to a correct downweighting level (root 1), as displayed in the middle panel of Figure 2, in the bottom left corner there are some solutions characterized by insufficient downweighting (root 2), whereas in the top right corner there are those solutions characterized by an excess of downweighting (root 3). Root 2 and root 3 are given in the middle and right panel, respectively, of Figure 3. The root selection strategy based on (5) correctly leads to choose root 1.

Example 2.

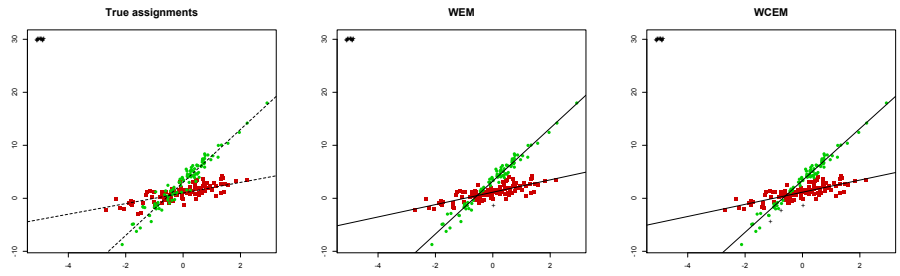

Fig. 4 Example 2. True assignments (left), WEM (middle), WCEM (right).

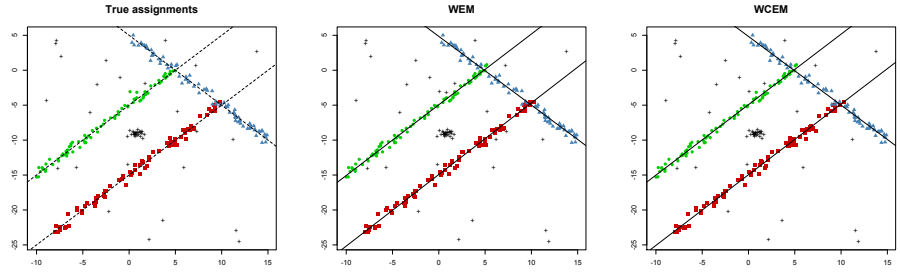

Fig. 5 Example 3. True assignments (left), WEM (middle), WCEM (right).

Let us consider a mixture of two regression lines. Genuine data are drawn according to the model

$$\begin{cases} y_1 = 1 + x + \epsilon \\ y_2 = 3 + 5x + \epsilon \end{cases}$$

with $\epsilon \sim N(0, 1)$ (Bai et al, 2012). Each group is composed by 100 points. By looking at the plots in Figure 4, we notice that the two clusters are overlapped and the regression lines share the same sign of the slope. Then, 20 clustered bad leverage points are added in the top left corner that violate the patterns exhibited by the genuine points. In this scenario, both the classical EM and CEM lead to a fitted mixture in which one fitted component is wrongly rotated and attracted by the outliers, whereas the other is not able to fit neither of the two true linear structures. On the contrary, the behavior of the robust techniques is satisfactory.

Example 3.

Let us consider a data constellation inspired by García-Escudero et al (2009). We have a mixture of three linear models disposed according to a slanted π configuration. The sample size is 300, data are simulated according to equal membership probabilities. There are 50 outliers that are of two types: 25 are scattered in the rectangle that contains the genuine observations, 25 are inliers, since they lie between the linear patterns. Figure 5 displays the data and the results. The weighted likelihood methodology still provides accurate and satisfactory results.

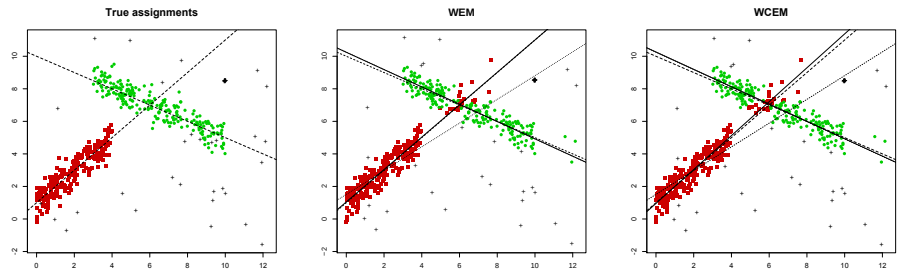

Fig. 6 Example 4. True assignments (left), WEM (middle), WCEM (right).

Example 4.

This example has been taken from García-Escudero et al (2010). In that paper, tha authors proposed TCLUST-REG allowing for a *second trimming* step to handle those data points acting as bad leverage points for the linear regressions. On the contrary, weighted likelihood regression is able to deal with outliers in the x-space and, according to our experience, there is not the need to introduce a second trimming. The data includes two linear regression clusters made up of 225 observations each from the model

$$\begin{cases} y_1 = 1 + x + 0.5\epsilon \\ y_2 = 10 - 0.5x + 0.5\epsilon \end{cases}$$

with $\epsilon \sim N(0,1)$. Then, 30 points are generated as a background noise and, finally, 20 more data points are concentrated around the point $(10, 8.5)$, acting as bad leverage points in the estimation of one linear structure. This data configuration will be also considered in the numerical studies in Section 6 as a part of larger numerical studies following the lines of García-Escudero et al (2010). Figure 6 displays the true assignments with the true lines and the fitted models by WEM and WCEM. In the middle and right panel, we superimposed both the true lines and the regression lines fitted by the trimmed likelihood, to better appreciate the nice behavior of WEM and WCEM in this scenario, since the trimmed likelihood approach of Neykov et al (2007) is not able to take into account bad leverages. Actually, trimmed likelihood estimation suffers from the presence of the group of bad leverages, since one regression line is rotated towards their direction. On the contrary, the weighted likelihood technique still gives robust estimates, in a fashion similar to TCLUST-REG, but without any second trimming. It is worth noting that both WEM and WCEM wrongly classify some data points, even if characterized by large uncertainties. Actually, the misclassified points by WEM and WCEM are about those trimmed in the second step of TCLUST-REG.

Example 5.

Here, we consider a data constellation similar to that analyzed in García-Escudero et al (2017) (see their Figure 6). As well as for TCWRM, the selection of appropriate restrictions on the variance regression error terms are needed in order to avoid undesired spurious solutions. The solution displayed in the middle

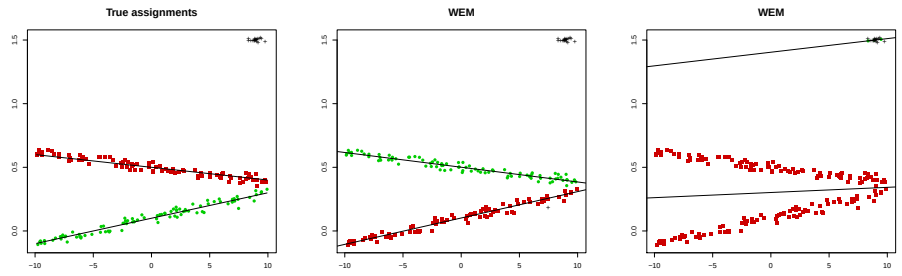

Fig. 7 Example 4. True assignments (left), WEM with $c = 5$ (middle), WEM with $c = 500$ (right).

panel of Figure 7 has been obtained for $c = 5$, whereas the one in the right panel corresponds to $c = 500$. Nevertheless, TCWRM needs the specification of a further constraint on the eigenvalues in the covariates' space. It is worth to notice that the root selection criterion based on the minimum fitted approximate disparity (5) leads to choose the right solution: we have $\tilde{\rho}(f^*, m^*) = 0.594$ for the good solution and $\tilde{\rho}(f^*, m^*) = 1.532$ for the spurious one.

6 Numerical studies

In this section the finite sample behavior of the proposed WEM and WCEM methodologies has been investigated by some numerical studies. The data generation scheme is similar to that proposed in García-Escudero et al (2010). We consider a mixture of two regression lines, i.e. with $p = 2$, according to the model described in Example 4. It is assumed that $x \sim U(D, D + 7)$, where the tuning parameter D controls the degree of overlapping by setting $D = 3, 6, 12$. Moreover, two different degrees of complexity have been taken into account: in the first we set equal clusters' proportions $\pi_1 = \pi_2$ and scales $\sigma_1 = \sigma_2 = 0.5$, whereas in the second we assumed unequal proportions and variances, with $\pi_1 = 0.6$, $\pi_2 = 0.4$, and $\sigma_1 = 0.4$, $\sigma_2 = 0.6$. The behavior of WEM and WCEM has been investigated both when any contamination does not occur ($\epsilon = 0$) and when outliers are present. For what concerns the contamination rates, we set $\epsilon = 10\%, 25\%$. Two types of outliers configurations have been considered. In the first scenario, outliers are generated as background noise (Cont.1), whereas in the second scenario we have both background noisy points and bad leverage points concentrated around a point mass (Cont.2). Then the numerical studies are composed by $2 \times 2 \times 3 \times (1 + 2 \times 2)$ separate simulations. The considered sample size is $n = 500$. Table 1 summarizes the structure of the data for each combination of complexity, scenario and outliers' rate. The numerical studies have been also carried out when the number of covariates is $p = 4$ by adding uninformative explanatory variables, that is the corresponding coefficients are set to zero.

The numerical studies are based on 500 Monte Carlo trials. The weighted likelihood algorithms are based on a symmetric Chi-square RAF. The smoothing parameter h has been selected in such a way that the empirical downweighting level lies in the range $(0.15, 0.20)$ for $\epsilon = 0.10$ and $(0.35, 0.45)$ for $\epsilon = 0.25$, whereas

it is about 0.10 when no outliers occur. The algorithm is assumed to reach convergence when $\max |\hat{\beta}^{(s+1)} - \hat{\beta}^{(s)}| < tol$, with a tolerance tol set to 10^{-4}, where $\hat{\beta}^{(s)}$ is the matrix of centroids estimates at the s^{th} iteration and the differences are elementwise. The algorithms run on non-optimized R code.

Fitting accuracy has been evaluated according to the Mean Squared Error (MSE) for the mixture parameters, whereas classification accuracy has been measured by the Adjusted Rand index (ARI) evaluated over true negatives, i.e. genuine observations that are not wrongly declared outliers. In order to detect outliers, we considered a testing rule with $\alpha = 0.01$, according to (7). In addition, we also adopted a strategy based on the FDR for the same overall level, in order to take into account multiplicity effects. Then, we reported the empirical level and power of the test, measured as the swamping rate and the rate of true positives as explained in Section 4. When many outliers are detected, than the power is expected to be high but genuine observations are likely to be misclassified, that is swamping also increases. On the other side, with a low rate of correctly flagged true outliers, the power and the level are expected to both decrease. Of course, when $\epsilon = 0$, swamping only is taken into account. The performance of the proposed WEM and WCEM has been compared with their M-type counterparts, MEM and MCEM, respectively, in which clusterwise M-estimation is performed at the M-step. Here, we considered M-estimation based on the Tukey biweight function for an 85% efficiency level.

As an overall result, we do appreciate the satisfactory behavior of all the four methods under investigation. The numerical studies did not unveil any remarkable difference between them both in terms of fitting and classification accuracy and for what concerns the task of outliers detection. The Tables that follow give detailed results. The entries in Table 2 give the ARI evaluated over true negatives after that outliers have been discarded according to a testing rule based on a fixed 1% level or by controlling the overall level of the multiple testing procedure by using the FDR. The results are quite satisfactory. The classification accuracy clearly improves for increasing values of the tuning parameter D and there are no relevant differences when using a fixed level or multiplicity issues are taken into account. Table 3 gives the MSE corresponding to the fitted mixture parameters (β, σ, π) stemming from all the considered techniques. The overall behavior of all the methods is quite accurate with weighted likelihood based methods leading often to smaller MSEs for all the parameters. Swamping and power of the outlier tests are given in Table 4 and Table 5, respectively. It is worth to stress that the behavior of the tests depends on the actual robustness-efficiency trade-off of the procedure, hence on the value of the selected bandwidth parameter h for weighted likelihood estimation. Here, we controlled the degree of robustness by selecting a different h for the two considered levels of contamination. Then, when $\epsilon = 10\%$ the weighted likelihood techniques are characterized by a lower rate of swamping and reasonable power with respect to their M-estimation based counterpart but the situation is reversed for $\epsilon = 25\%$. In the latter scenario, the power of the testing procedures stemming from WEM and WCEM is particularly appreciable. The entries in Tables from 6 to 9 give the results for the case $p = 4$. The finite sample behavior of the proposed methodologies is still accurate and satisfactory.

7 Pinus nigra data set

The following example has been taken from García-Escudero et al (2010). The data gives the height (in meters) and diameter (in millimeters) of $n = 362$ Pinus nigra trees located in the north of Palencia (Spain). The Diameter is considered as an explicative variables wheres Height is the response. The data are displayed in the left panel of Figure 8. They exhibit the presence of three linear groups apart from a small group of trees forming its own cluster on the top right corner and one isolated point on the bottom right corner. Therefore, we assume $k = 3$ and fit the model by WEM and WCEM, respectively, by setting $h = 0.01$ and employing a symmetric chi square RAF. The outlier detection rule is based on the FDR at 1% level. The fitted models and detected outliers are shown in Figure 8. The results are in strong agreement with those stemming from TCLUST-REG.

In order to explore more in details the procedure, let us look for possible multiple roots. Figure 9 gives the monitoring of the empirical downweighting level as h varies on a fixed grid of values for a couple of starting TCLUST solutions. We notice that two different solutions occur for $h < 0.014$, whereas for $h \geq 0.14$ both initializations lead to the same fitted model. The solutions denoted root2 are characterized by an excess of downweighting w.r.t the other solutions, named root1 for each $h < 0.014$. The abrupt shift in the root2 trajectory suggests a relevant structural change in the fitted model: for $h < 0.014$ the procedure gives place to many small weights and many outliers. On the contrary, the root1 trajectory does not suggest any substantial change in the fitted model. In order to examine more in depth the differences among the two solutions, let us take a look at the monitoring plots given in Figure 9 in the middle and right panels. Here, we monitor the change in individual residuals (in absolute value) as h varies for both considered initializations. The horizontal line in both panels gives the threshold for the outliers detection test at a fixed 0.01-level. The middle panel corresponds to the trajectory denoted root1 in the left panel, whereas the right panel to the other one named root2. The monitoring plot in the middle panel tells that the clustered outliers and the isolated outlier are clearly spotted during all the monitoring process and that the other data points have residuals below the threshold line for most of the monitoring. This means that the fitted model does not change remarkably as h varies and the genuine observations are assigned to a linear structure properly. The monitoring plot in the right panel tells a different story. Many trajectories are well above the chosen cut-off in its left hand section, that is many observations are downweighted and flagged as outliers. It is evident that, for values of the bandwidth parameter below a certain bound, there are at least two groups of outliers. In particular, the second (from top) group of outlying trajectories corresponds to a cluster of false positives rather than true outliers. Figure 10 displays the spurious solution obtained when $h = 0.01$: one genuine cluster is misclassified and its points are wrongly detected as outlying and the fitted mixture model is wrong. In particular, we notice that two components are wrongly fitted since the group in the middle has been erroneously split.

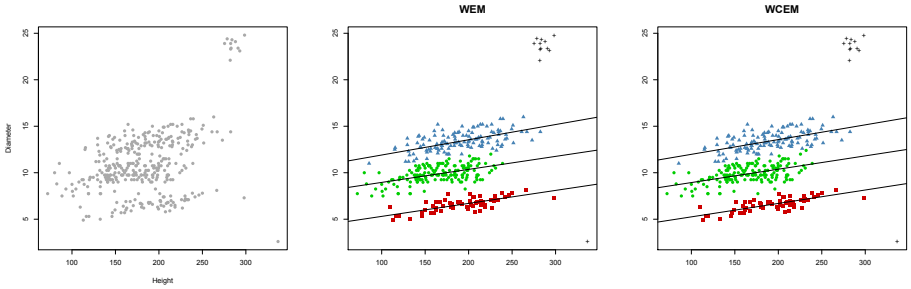

Fig. 8 Pinus nigra. Original data (left). Fitted mixtures by WEM (middle) and WCEM (right). Outliers are denoted by +.

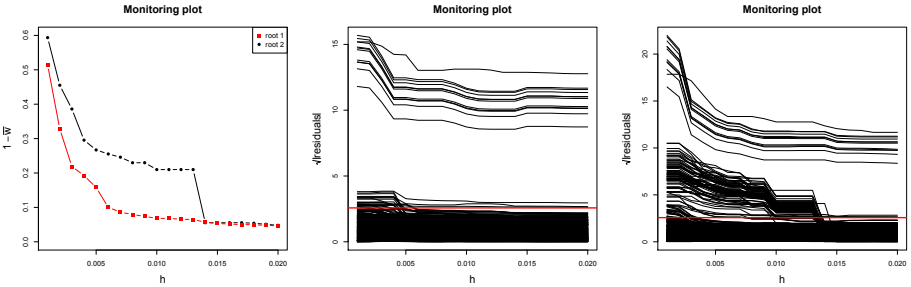

Fig. 9 Pinus nigra. Monitoring of the empirical downweighting level from WEM by using two different starting TCLUST solutions (left). Monitoring of clusterwise residuals in absolute value from WEM initialized by the two different starting TCLUST solutions: root1 (middle), root2 (right).

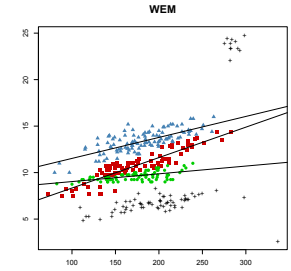

Fig. 10 Pinus nigra. Spurious root by WEM: classification and fitted model after outlier detection. Outliers are denoted by +.

References

Agostinelli C (2002) Robust model selection in regression via weighted likelihood methodology. Statistics & probability letters 56(3):289–300

Agostinelli C (2006) Notes on pearson residuals and weighted likelihood estimating equations. Statistics & probability letters 76(17):1930–1934

Agostinelli C, Greco L (2013) A weighted strategy to handle likelihood uncertainty in bayesian inference. Computational Statistics 28(1):319–339

Agostinelli C, Greco L (2017) Discussion on "the power of monitoring: How to make the most of a contaminated sample". Statistical Methods & Applications 27(4):609–619

Agostinelli C, Greco L (2019) Weighted likelihood estimation of multivariate location and scatter. Test 28(3):756–784

Agostinelli C, Markatou M (1998) A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. Statistics & probability letters 37(4):341–350

Agostinelli C, Markatou M (2001) Test of hypotheses based on the weighted likelihood methodology. Statistica Sinica pp 499–514

Alqallaf F, Agostinelli C (2016) Robust inference in generalized linear models. Communications in Statistics-Simulation and Computation 45(9):3053–3073

Bai X, Yao W, Boyer JE (2012) Robust fitting of mixture regression models. Computational Statistics & Data Analysis 56(7):2347–2359

Bashir S, Carter E (2012) Robust mixture of linear regression models. Communications in Statistics-Theory and Methods 41(18):3371–3388

Basu A, Lindsay B (1994) Minimum disparity estimation for continuous models: efficiency, distributions and robustness. Annals of the Institute of Statistical Mathematics 46(4):683–705

Campbell N (1984) Mixture models and atypical values. Mathematical Geology 16(5):465–477

Cerioli A, Farcomeni A (2011) Error rates for multivariate outlier detection. Computational Statistics & Data Analysis 55(1):544–553

Cerioli A, Riani M, Atkinson AC, Corbellini A (2018) The power of monitoring: how to make the most of a contaminated multivariate sample. Statistical Methods & Applications pp 1–29

Coretto P, Hennig C (2017) Consistency, breakdown robustness, and algorithms for robust improper maximum likelihood clustering. The Journal of Machine Learning Research 18(1):5199–5237

Dotto F, Farcomeni A, García-Escudero LA, Mayo-Iscar A (2017) A fuzzy approach to robust regression clustering. Advances in Data Analysis and Classification 11(4):691–710

Elashoff M, Ryan L (2004) An em algorithm for estimating equations. Journal of Computational and Graphical Statistics 13(1):48–65

Farcomeni A, Greco L (2015a) Robust methods for data reduction. CRC press

Farcomeni A, Greco L (2015b) S-estimation of hidden markov models. Computational Statistics 30(1):57–80

Fritz H, Garcia-Escudero L, Mayo-Iscar A (2013) A fast algorithm for robust constrained clustering. Computational Statistics & Data Analysis 61:124–136

García-Escudero L, Gordaliza A, Matrán C, Mayo-Iscar A (2008) A general trimming approach to robust cluster analysis. The Annals of Statistics 36(3):1324–1345

Garcia-Escudero L, Gordaliza A, Matran C, Mayo-Iscar A (2015) Avoiding spurious local maximizers in mixture modeling. Statistics and Computing 25(3):619–633

García-Escudero LA, Gordaliza A, San Martin R, Van Aelst S, Zamar R (2009) Robust linear clustering. Journal of the Royal Statistical Society: Series B (Sta-

tistical Methodology) 71(1):301–318
García-Escudero LA, Gordaliza A, Mayo-Iscar A, San Martín R (2010) Robust clusterwise linear regression through trimming. Computational Statistics & Data Analysis 54(12):3057–3069
García-Escudero LA, Gordaliza A, Greselin F, Ingrassia S, Mayo-Íscar A (2017) Robust estimation of mixtures of regressions with random covariates, via trimming and constraints. Statistics and Computing 27(2):377–402
Greco L (2017) Weighted likelihood based inference for $p(x < y)$. Communications in Statistics-Simulation and Computation 46(10):7777–7789
Greco L, Agostinelli C (2019) Weighted likelihood mixture modeling and model-based clustering. Statistics and Computing https://doi.org/10.1007/s11222-019-09881-1
Hennig C (2000) Identifiablity of models for clusterwise linear regression. Journal of Classification 17(2):273–296
Markatou M, Basu A, Lindsay BG (1998) Weighted likelihood equations with bootstrap root search. Journal of the American Statistical Association 93(442):740–750
Maronna RA, Martin RD, Yohai VJ, Salibián-Barrera M (2018) Robust statistics: theory and methods (with R). Wiley
Neykov N, Filzmoser P, Dimova R, Neytchev P (2007) Robust fitting of mixtures using the trimmed likelihood estimator. Computational Statistics & Data Analysis 52(1):299–308
Neykov NM, Müller CH (2003) Breakdown point and computation of trimmed likelihood estimators in generalized linear models. In: Developments in robust statistics, Springer, pp 277–286
Punzo A, McNicholas PD (2017) Robust clustering in regression analysis via the contaminated gaussian cluster-weighted model. Journal of Classification 34(2):249–293
Riani M, Cerioli A, Atkinson A, Perrotta D, Torti F (2008) Fitting mixtures of regression lines with the forward search. Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security 19:271
Torti F, Perrotta D, Riani M, Cerioli A (2019) Assessing trimming methodologies for clustering linear regression data. Advances in Data Analysis and Classification 13(1):227–257
Yao W, Wei Y, Yu C (2014) Robust mixture regression using the t-distribution. Computational Statistics & Data Analysis 71:116–127
Yu C, Yao W, Chen K (2017) A new method for robust mixture regression. Canadian Journal of Statistics 45(1):77–94

Table 1 Data configurations used in the numerical studies with $n = 500$.

Complexity	Scenario	ϵ	background outliers	bad leverages
	No contamination	0	0	0
$\pi_1 = \pi_2$	Cont.1	0.10	50	0
$\sigma_1 = \sigma_2$		0.25	125	0
	Cont.2	0.10	30	20
		0.25	75	50
	No contamination	0	0	0
$\pi_1 \neq \pi_2$	Cont.1	0.10	50	0
$\sigma_1 \neq \sigma_2$		0.25	125	0
	Cont.2	0.10	30	20
		0.25	75	50

Table 2 Adjusted Rand Index evaluated over true negatives for WEM, WCEM, MEM and MCEM, $p = 2$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters. The outlier detection rule is based on a 0.01 level by using a fixed level and the False Discovery Rate.

		WEM		WCEM		MEM		MCEM	
		Fixed	FDR	Fixed	FDR	Fixed	FDR	Fixed	FDR
		$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$							
No Cont	D=3	0.801	0.803	0.797	0.799	0.807	0.812	0.815	0.815
	D=6	0.876	0.877	0.889	0.890	0.883	0.886	0.897	0.897
	D=12	0.999	0.999	0.999	0.999	1.000	0.999	0.999	0.999
Cont.1 10%	D=3	0.809	0.811	0.808	0.811	0.812	0.817	0.816	0.821
	D=6	0.874	0.876	0.892	0.893	0.878	0.879	0.894	0.895
	D=12	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Cont.1 25%	D=3	0.792	0.794	0.782	0.783	0.792	0.795	0.799	0.802
	D=6	0.863	0.865	0.873	0.875	0.855	0.857	0.875	0.877
	D=12	0.997	0.998	0.996	0.997	0.996	0.997	0.997	0.998
Cont.2 10%	D=3	0.810	0.812	0.801	0.803	0.809	0.813	0.810	0.814
	D=6	0.871	0.873	0.882	0.884	0.881	0.884	0.894	0.896
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.2 25%	D=3	0.796	0.799	0.777	0.799	0.810	0.813	0.811	0.815
	D=6	0.849	0.862	0.871	0.873	0.869	0.872	0.896	0.897
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$							
No cont	D=3	0.806	0.809	0.808	0.811	0.813	0.818	0.823	0.823
	D=6	0.889	0.890	0.900	0.901	0.895	0.898	0.906	0.906
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.1 10%	D=3	0.807	0.810	0.821	0.824	0.814	0.818	0.821	0.825
	D=6	0.883	0.884	0.905	0.907	0.890	0.892	0.898	0.900
	D=12	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
Cont.1 25%	D=3	0.802	0.812	0.842	0.845	0.804	0.806	0.814	0.817
	D=6	0.883	0.890	0.918	0.920	0.883	0.884	0.899	0.900
	D=12	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
Cont.2 10%	D=3	0.814	0.816	0.824	0.826	0.817	0.820	0.824	0.827
	D=6	0.887	0.888	0.903	0.904	0.894	0.895	0.901	0.902
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.2 25%	D=3	0.819	0.823	0.851	0.854	0.817	0.820	0.825	0.828
	D=6	0.887	0.889	0.915	0.917	0.885	0.886	0.900	0.902
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3 Mean Squared Error for WEM, WCEM, MEM and MCEM, $p = 2$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters.

	WEM			WCEM			MEM			MCEM		
	β	σ	π	β	σ	π	β	σ	π	β	σ	π
	$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$											
	No cont											
D=3	0.024	0.010	0.006	0.025	0.010	0.007	0.032	0.048	0.006	0.036	0.054	0.006
D=6	0.088	0.008	0.002	0.116	0.008	0.002	0.096	0.044	0.002	0.116	0.048	0.002
D=12	0.080	0.010	0.001	0.080	0.006	0.001	0.104	0.042	0.001	0.105	0.041	0.001
	Cont.1 - 10 %											
D=3	0.039	0.015	0.006	0.041	0.026	0.010	0.043	0.047	0.007	0.042	0.052	0.007
D=6	0.107	0.015	0.003	0.126	0.024	0.004	0.099	0.046	0.003	0.118	0.049	0.003
D=12	0.112	0.014	0.001	0.109	0.020	0.002	0.112	0.044	0.002	0.112	0.044	0.002
	Cont.1 - 25 %											
D=3	0.092	0.058	0.009	0.092	0.071	0.019	0.100	0.048	0.008	0.083	0.095	0.008
D=6	0.312	0.055	0.004	0.345	0.063	0.011	0.376	0.044	0.005	0.389	0.048	0.004
D=12	0.592	0.051	0.002	0.639	0.055	0.003	0.736	0.042	0.001	0.679	0.042	0.001
	Cont.2 - 10 %											
D=3	0.038	0.019	0.007	0.036	0.032	0.015	0.039	0.051	0.008	0.040	0.057	0.008
D=6	0.115	0.018	0.004	0.129	0.024	0.007	0.121	0.053	0.006	0.131	0.053	0.004
D=12	0.088	0.015	0.001	0.092	0.019	0.002	0.121	0.048	0.002	0.121	0.048	0.002
	Cont.2 - 25 %											
D=3	0.101	0.036	0.008	0.097	0.062	0.015	0.101	0.065	0.012	0.083	0.065	0.010
D=6	0.230	0.040	0.009	0.243	0.050	0.014	0.234	0.066	0.012	0.211	0.056	0.007
D=12	0.231	0.052	0.002	0.253	0.055	0.005	0.259	0.047	0.004	0.259	0.047	0.004
	$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$											
	No cont											
D=3	0.027	0.013	0.006	0.028	0.010	0.004	0.035	0.049	0.006	0.039	0.053	0.004
D=6	0.087	0.012	0.002	0.103	0.020	0.002	0.101	0.046	0.002	0.115	0.050	0.003
D=12	0.092	0.010	0.001	0.089	0.006	0.000	0.124	0.044	0.001	0.124	0.044	0.001
	Cont.1 - 10%											
D=3	0.041	0.017	0.007	0.039	0.022	0.004	0.046	0.050	0.006	0.044	0.055	0.006
D=6	0.151	0.016	0.003	0.165	0.020	0.002	0.128	0.047	0.003	0.155	0.050	0.003
D=12	0.123	0.014	0.001	0.134	0.016	0.003	0.111	0.043	0.001	0.111	0.043	0.001
	Cont.1 - 25 %											
D=3	0.071	0.061	0.007	0.067	0.060	0.011	0.018	0.051	0.007	0.070	0.056	0.007
D=6	0.256	0.058	0.004	0.270	0.063	0.014	0.343	0.047	0.004	0.324	0.051	0.003
D=12	0.521	0.054	0.019	0.885	0.044	0.047	0.845	0.045	0.038	0.707	0.045	0.039
	Cont.2 - 10%											
D=3	0.030	0.020	0.007	0.031	0.026	0.006	0.037	0.052	0.006	0.037	0.057	0.006
D=6	0.106	0.019	0.004	0.123	0.023	0.003	0.119	0.051	0.004	0.133	0.053	0.003
D=12	0.098	0.017	0.001	0.097	0.019	0.003	0.127	0.047	0.001	0.127	0.047	0.001
	Cont.2 - 25 %											
D=3	0.071	0.064	0.008	0.060	0.069	0.013	0.059	0.053	0.008	0.059	0.056	0.006
D=6	0.192	0.059	0.004	0.176	0.065	0.013	0.118	0.049	0.003	0.200	0.052	0.003
D=12	0.275	0.055	0.002	0.301	0.060	0.020	0.331	0.046	0.002	0.331	0.046	0.002

Table 4 Swamping rate for WEM, WCEM, MEM and MCEM, $p = 2$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters. The outlier detection rule is based on a 0.01 level by using a fixed level and the False Discovery Rate.

		WEM		WCEM		MEM		MCEM	
		Fixed	FDR	Fixed	FDR	Fixed	FDR1	Fixed	FDR
		$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$							
No Cont	D=3	0.022	0.000	0.022	0.000	0.066	0.008	0.000	0.000
	D=6	0.022	0.000	0.021	0.000	0.068	0.008	0.000	0.000
	D=12	0.026	0.000	0.021	0.000	0.067	0.008	0.000	0.000
Cont.1 - 10%	D=3	0.030	0.008	0.043	0.014	0.068	0.027	0.076	0.031
	D=6	0.032	0.007	0.041	0.011	0.068	0.026	0.073	0.029
	D=12	0.033	0.008	0.040	0.011	0.072	0.027	0.072	0.027
Cont.1 - 25%	D=3	0.052	0.050	0.075	0.063	0.067	0.038	0.076	0.044
	D=6	0.071	0.049	0.072	0.057	0.066	0.037	0.071	0.041
	D=12	0.078	0.047	0.069	0.052	0.066	0.037	0.066	0.037
Cont.2 - 10%	D=3	0.034	0.009	0.047	0.015	0.073	0.030	0.081	0.035
	D=6	0.036	0.009	0.041	0.011	0.082	0.039	0.079	0.035
	D=12	0.034	0.008	0.038	0.010	0.076	0.033	0.076	0.033
Cont.2 - 25%	D=3	0.058	0.031	0.087	0.056	0.101	0.071	0.095	0.063
	D=6	0.079	0.045	0.077	0.045	0.109	0.075	0.087	0.057
	D=12	0.081	0.047	0.085	0.051	0.075	0.044	0.087	0.057
		$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$							
No cont	D=3	0.025	0.000	0.023	0.000	0.068	0.008	0.000	0.000
	D=6	0.025	0.000	0.022	0.000	0.066	0.008	0.000	0.000
	D=12	0.026	0.001	0.021	0.000	0.067	0.007	0.000	0.000
Cont.1 - 10%	D=3	0.031	0.007	0.038	0.011	0.069	0.027	0.076	0.032
	D=6	0.032	0.007	0.038	0.010	0.068	0.027	0.072	0.029
	D=12	0.030	0.007	0.034	0.009	0.067	0.024	0.068	0.024
Cont.1 - 25%	D=3	0.083	0.049	0.104	0.070	0.068	0.038	0.076	0.044
	D=6	0.085	0.051	0.103	0.070	0.070	0.040	0.074	0.042
	D=12	0.083	0.049	0.078	0.047	0.069	0.038	0.069	0.039
Cont.2 - 10%	D=3	0.033	0.009	0.041	0.013	0.070	0.027	0.077	0.032
	D=6	0.034	0.009	0.040	0.012	0.072	0.030	0.075	0.030
	D=12	0.034	0.009	0.039	0.011	0.071	0.028	0.072	0.028
Cont.2 - 25%	D=3	0.090	0.056	0.104	0.070	0.074	0.044	0.077	0.045
	D=6	0.084	0.050	0.102	0.067	0.068	0.038	0.072	0.041
	D=12	0.084	0.049	0.104	0.070	0.070	0.040	0.071	0.040

Table 5 Power of the outlier test for WEM, WCEM, MEM and MCEM, $p = 2$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters. The outlier detection rule is based on a 0.01 level by using a fixed level and the False Discovery Rate.

		WEM		WCEM		MEM		MCEM	
		Fixed	FDR	Fixed	FDR	Fixed	FDR	Fixed	FDR
		$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$							
Cont.1 - 10%	D=3	0.981	0.912	0.988	0.938	0.993	0.975	0.995	0.980
	D=6	0.976	0.919	0.988	0.941	0.990	0.970	0.994	0.976
	D=12	0.988	0.928	0.991	0.944	0.998	0.981	0.998	0.981
Cont.1 - 25%	D=3	0.968	0.956	0.966	0.953	0.971	0.958	0.982	0.971
	D=6	0.978	0.969	0.982	0.973	0.972	0.959	0.978	0.968
	D=12	0.986	0.982	0.985	0.979	0.983	0.974	0.985	0.975
Cont.2 - 10%	D=3	0.962	0.962	0.989	0.989	0.984	0.984	0.984	0.984
	D=6	0.918	0.918	0.962	0.962	0.957	0.957	0.978	0.978
	D=12	0.989	0.989	0.995	0.995	0.979	0.979	0.979	0.979
Cont.2 - 25%	D=3	0.927	0.913	0.950	0.943	0.944	0.939	0.963	0.959
	D=6	0.926	0.923	0.947	0.940	0.924	0.915	0.969	0.963
	D=12	0.997	0.993	0.996	0.993	0.980	0.980	0.986	0.980
		$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$							
Cont.1 - 10%	D=3	0.978	0.915	0.985	0.930	0.992	0.972	0.995	0.978
	D=6	0.972	0.913	0.980	0.925	0.990	0.966	0.993	0.972
	D=12	0.988	0.923	0.987	0.923	0.998	0.979	0.998	0.979
Cont.1 - 25%	D=3	0.987	0.980	0.991	0.985	0.980	0.972	0.987	0.980
	D=6	0.987	0.979	0.987	0.979	0.977	0.967	0.983	0.976
	D=12	0.990	0.984	0.977	0.963	0.983	0.975	0.984	0.976
Cont.2 - 10%	D=3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	D=6	0.984	0.984	0.995	0.995	0.995	0.995	1.000	1.000
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.2 - 25%	D=3	0.986	0.982	0.993	0.988	0.982	0.977	0.992	0.987
	D=6	0.992	0.987	0.994	0.989	0.992	0.985	0.994	0.989
	D=12	0.997	0.993	0.995	0.990	0.991	0.987	0.991	0.987

Table 6 Adjusted Rand Index evaluated over true negatives for WEM, WCEM, MEM and MCEM, $p = 4$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters. The outlier detection rule is based on a 0.01 level by using a fixed level and the False Discovery Rate.

		WEM		WCEM		MEM		MCEM	
		Fix01	Fdr01	Fix01	Fdr01	Fix01	Fdr01	Fix01	Fdr01
		$n_1 = n_2$ and $\sigma_1 = \sigma_2$							
No Cont	D=3	0.809	0.811	0.805	0.807	0.813	0.817	0.815	0.819
	D=6	0.874	0.876	0.887	0.888	0.880	0.884	0.893	0.896
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.1 10%	D=3	0.807	0.809	0.809	0.812	0.812	0.815	0.816	0.819
	D=6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.1 25%	D=3	0.789	0.791	0.778	0.779	0.793	0.794	0.801	0.804
	D=6	0.857	0.859	0.871	0.872	0.855	0.857	0.877	0.880
	D=12	0.997	0.997	0.997	0.996	0.994	0.994	0.998	0.998
Cont.2 10%	D=3	0.807	0.809	0.802	0.805	0.815	0.819	0.817	0.821
	D=6	0.860	0.862	0.878	0.880	0.883	0.886	0.895	0.897
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.2 25%	D=3	0.804	0.807	0.770	0.771	0.803	0.806	0.809	0.814
	D=6	0.870	0.873	0.863	0.865	0.883	0.885	0.899	0.900
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$n_1 \neq n_2$ and $\sigma_1 \neq \sigma_2$							
No cont	D=3	0.808	0.811	0.815	0.818	0.817	0.823	0.821	0.828
	D=6	0.884	0.885	0.893	0.895	0.887	0.890	0.895	0.897
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.1 10%	D=3	0.803	0.806	0.824	0.826	0.812	0.817	0.818	0.823
	D=6	0.887	0.887	0.906	0.907	0.892	0.895	0.904	0.906
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.1 25%	D=3	0.803	0.806	0.838	0.841	0.804	0.808	0.814	0.818
	D=6	0.877	0.879	0.914	0.915	0.880	0.882	0.897	0.899
	D=12	0.993	0.993	0.988	0.988	0.993	0.993	0.993	0.993
Cont.2 10%	D=3	0.815	0.817	0.828	0.829	0.820	0.822	0.827	0.830
	D=6	0.886	0.887	0.907	0.908	0.896	0.898	0.903	0.905
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.2 25%	D=3	0.816	0.819	0.833	0.837	0.819	0.822	0.823	0.827
	D=6	0.890	0.892	0.907	0.909	0.892	0.894	0.898	0.900
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 7 Mean Squared Error for WEM, WCEM, MEM and MCEM, $p = 4$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters.

	WEM			WCEM			MEM			MCEM		
	β	σ	π	β	σ	π	β	σ	π	β	σ	π
	$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$											
	No cont											
D=3	0.053	0.012	0.004	0.048	0.010	0.003	0.086	0.052	0.006	0.076	0.058	0.006
D=6	0.087	0.012	0.003	0.097	0.009	0.001	0.102	0.050	0.002	0.112	0.047	0.002
D=12	0.081	0.011	0.001	0.077	0.006	0.001	0.101	0.045	0.001	0.103	0.046	0.001
	Cont.1 - 10%											
D=3	0.094	0.013	0.007	0.096	0.021	0.009	0.119	0.053	0.008	0.119	0.059	0.007
D=6	0.245	0.011	0.001	0.231	0.015	0.002	0.190	0.048	0.001	0.190	0.048	0.001
D=12	0.174	0.011	0.001	0.171	0.016	0.002	0.166	0.049	0.002	0.166	0.049	0.002
	Cont.1 - 25%											
D=3	0.203	0.045	0.009	0.202	0.061	0.027	0.200	0.053	0.010	0.185	0.061	0.009
D=6	0.485	0.043	0.005	0.514	0.052	0.008	0.477	0.051	0.005	0.472	0.056	0.004
D=12	0.775	0.035	0.002	0.825	0.044	0.006	0.959	0.047	0.002	0.773	0.047	0.002
	Cont.2 - 10%											
D=3	0.089	0.016	0.007	0.083	0.027	0.014	0.104	0.057	0.008	0.106	0.063	0.008
D=6	0.188	0.015	0.006	0.190	0.023	0.008	0.181	0.061	0.007	0.200	0.063	0.006
D=12	0.139	0.013	0.001	0.145	0.018	0.002	0.190	0.055	0.004	0.190	0.055	0.004
	Cont.2 - 25%											
D=3	0.230	0.065	0.013	0.242	0.066	0.029	0.223	0.072	0.013	0.203	0.067	0.009
D=6	0.343	0.070	0.012	0.339	0.063	0.027	0.325	0.087	0.020	0.306	0.069	0.009
D=12	0.346	0.042	0.002	0.345	0.047	0.005	0.377	0.057	0.007	0.377	0.057	0.007
	$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$											
	No cont											
D=3	0.056	0.014	0.006	0.056	0.010	0.004	0.076	0.052	0.004	0.080	0.058	0.004
D=6	0.087	0.013	0.002	0.085	0.010	0.002	0.102	0.052	0.002	0.112	0.055	0.002
D=12	0.084	0.010	0.001	0.080	0.006	0.001	0.101	0.046	0.001	0.108	0.046	0.001
	Cont.1 - 10%											
D=3	0.098	0.013	0.007	0.099	0.020	0.005	0.117	0.054	0.007	0.115	0.059	0.007
D=6	0.211	0.012	0.003	0.235	0.018	0.002	0.208	0.051	0.003	0.229	0.054	0.002
D=12	0.190	0.012	0.001	0.197	0.016	0.003	0.206	0.050	0.002	0.206	0.050	0.002
	Cont.1 - 25%											
D=3	0.172	0.049	0.007	0.177	0.058	0.008	0.177	0.058	0.007	0.171	0.064	0.006
D=6	0.453	0.045	0.004	0.478	0.053	0.010	0.437	0.053	0.004	0.427	0.057	0.004
D=12	0.657	0.043	0.002	0.975	0.052	0.022	0.675	0.002	0.671	0.051	0.002	
	Cont.2 - 10%											
D=3	0.085	0.017	0.007	0.089	0.027	0.006	0.112	0.060	0.007	0.114	0.066	0.007
D=6	0.168	0.015	0.004	0.176	0.023	0.003	0.200	0.059	0.005	0.210	0.060	0.004
D=12	0.145	0.014	0.001	0.150	0.019	0.003	0.197	0.053	0.002	0.197	0.053	0.002
	Cont.2 - 25%											
D=3	0.111	0.050	0.007	0.120	0.058	0.007	0.109	0.056	0.007	0.112	0.061	0.006
D=6	0.189	0.049	0.004	0.201	0.055	0.007	0.187	0.054	0.004	0.205	0.057	0.003
D=12	0.225	0.045	0.001	0.217	0.048	0.008	0.229	0.053	0.002	0.225	0.052	0.002

Table 8 Swamping rate for WEM, WCEM, MEM and MCEM, $p = 4$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters. The outlier detection rule is based on a 0.01 level by using a fixed level and the False Discovery Rate.

		WEM		WCEM		MEM		MCEM	
		Fixed	FDR	Fixed	FDR	Fixed	FDR	Fixed	FDR
		$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$							
No Cont	D=3	0.026	0.000	0.023	0.000	0.072	0.012	0.081	0.017
	D=6	0.026	0.000	0.022	0.000	0.073	0.011	0.079	0.013
	D=12	0.026	0.000	0.021	0.000	0.073	0.010	0.073	0.010
Cont.1 10%	D=3	0.027	0.007	0.037	0.011	0.076	0.033	0.084	0.038
	D=6	0.029	0.007	0.034	0.009	0.076	0.032	0.076	0.032
	D=12	0.037	0.010	0.031	0.007	0.075	0.031	0.076	0.030
Cont.1 25%	D=3	0.066	0.038	0.084	0.054	0.075	0.045	0.087	0.054
	D=6	0.069	0.039	0.079	0.048	0.078	0.046	0.083	0.052
	D=12	0.066	0.036	0.073	0.042	0.076	0.044	0.076	0.044
Cont.2 - 10%	D=3	0.031	0.008	0.043	0.013	0.081	0.036	0.091	0.042
	D=6	0.035	0.008	0.042	0.012	0.096	0.050	0.097	0.051
	D=12	0.032	0.007	0.038	0.011	0.090	0.045	0.090	0.045
Cont.2 - 25%	D=3	0.107	0.077	0.098	0.062	0.116	0.083	0.102	0.067
	D=6	0.122	0.091	0.098	0.065	0.147	0.118	0.110	0.079
	D=12	0.070	0.039	0.077	0.045	0.094	0.063	0.094	0.063
		$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$							
No cont	D=3	0.033	0.007	0.026	0.005	0.076	0.031	0.085	0.039
	D=6	0.038	0.009	0.026	0.005	0.075	0.030	0.080	0.033
	D=12	0.028	0.006	0.024	0.005	0.071	0.027	0.071	0.027
Cont.1 - 10%	D=3	0.027	0.006	0.037	0.010	0.076	0.033	0.085	0.039
	D=6	0.027	0.006	0.036	0.010	0.073	0.032	0.076	0.034
	D=12	0.028	0.007	0.036	0.010	0.077	0.032	0.077	0.032
Cont.1 - 25%	D=3	0.069	0.039	0.092	0.062	0.078	0.048	0.088	0.056
	D=6	0.070	0.039	0.090	0.059	0.079	0.047	0.084	0.050
	D=12	0.069	0.039	0.096	0.065	0.079	0.047	0.079	0.047
Cont.2 - 10%	D=3	0.030	0.008	0.044	0.015	0.084	0.039	0.091	0.045
	D=6	0.029	0.006	0.026	0.005	0.071	0.028	0.078	0.032
	D=12	0.027	0.006	0.025	0.005	0.072	0.027	0.073	0.027
Cont.2 - 25%	D=3	0.070	0.028	0.086	0.041	0.077	0.033	0.083	0.038
	D=6	0.075	0.033	0.085	0.040	0.085	0.035	0.078	0.035
	D=12	0.071	0.028	0.082	0.037	0.083	0.037	0.080	0.034

Table 9 Power of the outlier test for WEM, WCEM, MEM and MCEM, $p = 4$, for different type of contamination, rate of contamination and degree of overlapping among linear clusters. The outlier detection rule is based on a 0.01 level by using a fixed level and the False Discovery Rate.

		WEM		WCEM		MEM		MCEM	
		Fix01	Fdr01	Fix01	Fdr01	Fix01	Fdr01	Fix01	Fdr01
		$\pi_1 = \pi_2$ and $\sigma_1 = \sigma_2$							
Cont.1 - 10%	D=3	0.973	0.900	0.980	0.927	0.993	0.975	0.995	0.980
	D=6	0.978	0.916	0.982	0.929	0.997	0.982	0.997	0.982
	D=12	0.988	0.927	0.981	0.909	0.997	0.978	0.997	0.999
Cont.1 - 25%	D=3	0.971	0.957	0.974	0.963	0.973	0.964	0.982	0.975
	D=6	0.966	0.952	0.970	0.959	0.969	0.959	0.977	0.967
	D=12	0.982	0.971	0.981	0.968	0.981	0.974	0.983	0.976
Cont.2 - 10%	D=3	0.956	0.956	1.000	1.000	1.000	1.000	1.000	1.000
	D=6	0.824	0.824	0.941	0.941	0.954	0.954	0.965	0.965
	D=12	0.968	0.968	0.979	0.979	0.980	0.980	0.980	0.980
Cont.2 - 25%	D=3	0.914	0.897	0.921	0.914	0.916	0.910	0.958	0.953
	D=6	0.968	0.954	0.974	0.962	0.970	0.960	0.978	0.968
	D=12	0.991	0.985	0.992	0.986	0.979	0.975	0.979	0.975
		$\pi_1 \neq \pi_2$ and $\sigma_1 \neq \sigma_2$							
Cont.1 - 10%	D=3	0.973	0.900	0.981	0.923	0.991	0.971	0.994	0.978
	D=6	0.968	0.894	0.976	0.916	0.988	0.968	0.992	0.971
	D=12	0.982	0.921	0.983	0.929	0.999	0.983	0.999	0.983
Cont.1 - 25%	D=3	0.975	0.972	0.979	0.978	0.977	0.974	0.982	0.982
	D=6	0.975	0.963	0.979	0.966	0.977	0.968	0.982	0.975
	D=12	0.981	0.970	0.973	0.960	0.981	0.974	0.981	0.974
Cont.2 - 10%	D=3	0.989	0.989	0.989	0.989	0.989	0.989	0.989	0.989
	D=6	0.946	0.894	0.954	0.901	0.988	0.963	0.988	0.965
	D=12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cont.2 - 25%	D=3	0.996	0.996	1.000	1.000	0.996	0.996	1.000	1.000
	D=6	0.986	0.985	0.996	0.996	0.995	0.995	1.000	1.000
	D=12	1.000	1.000	1.000	1.000	0.991	0.990	1.000	1.000